The *erg* gene: A human gene related to the *ets* oncogene

(cloning/sequencing/expression/oncogene homology)

E. Shyam P. Reddy*†, Veena N. Rao‡, and Takis S. Papas‡

‡Laboratory of Molecular Oncology and *Programme Resources, Inc., National Cancer Institute, Frederick, MD 21701

ABSTRACT We have isolated a cDNA clone representing the complete coding sequence of a human gene named *erg*, related to the *ets* oncogene. Nucleotide sequence analysis of this cDNA (4.6 kilobases long) revealed that this gene encodes a 363-residue protein whose predicted amino acid sequence showed a homology of ≈40% and ≈70% to two domains corresponding to the 5′ and 3′ regions of v-*ets* oncogene, respectively. A 3.2- to 3.6-kilobase and ≈5-kilobase transcript of the *erg* gene, which differ in size from those of the previously described Hu-*ets* 1 and Hu-*ets* 2 genes, were observed in different cells. These results suggest that the *erg* gene is a member of the *ets* oncogene family.

E26 is a replication-defective avian acute leukemia virus, which causes erythroblastosis and a low level of concomitant myeloblastosis in chickens (1, 2). The genome of E26 contains, in addition to the v-*myb* oncogene, a second cell-derived oncogene v-*ets* (3, 4). The v-*myb* and v-*ets* oncogenes are expressed, together with a truncated viral *gag* gene, as a 135-kDa (p135) protein (3). Recently, it was shown that there exist two *ets* loci in humans, Hu-*ets1* and Hu-*ets2*, which were mapped to human chromosomes 11 and 21, respectively (5, 6).§ The Hu-*ets1* encodes a single mRNA of 6.8 kilobases (kb) and Hu-*ets2* encodes three mRNAs of 4.7, 3.2, and 2.7 kb (5). By contrast, the chicken homologue has contiguous *ets1* and *ets2* sequences and is expressed in normal chicken cells as a single 7.5-kb mRNA (4, 5).

Normal cellular genes that have given rise to retroviral transforming sequences represent an abundant class of genes. In addition to these genes, dominant transforming genes present in certain tumor cells have been found. Recently, other genes have been implicated in the transformation process by virtue of their amplification in tumor cells and they are found to be related to known oncogenes—for example, N-*myc* (8, 9), L-*myc* (10), c-*erb* B2 (11), and *neu* (12). In addition, other genes related to oncogenes—namely, *arg* (13), *slk/syn* (14, 15), *pks* (16), *ral* (17), and *rho* (18)—have been described.

In this study, we have made a cDNA library from COLO 320 cells and have isolated cDNA clones representing the complete coding sequence of an *ets*-related gene named *erg*. We have characterized the cDNA clones by sequencing and studied their expression in various cells. The longest cDNA clone, λ7 (Fig. 1), was chosen for further characterization.

MATERIALS AND METHODS

Cells. COLO 320 cells were maintained in RPMI 1640 medium with 10% calf serum.

cDNA Library. A cDNA library from human COLO 320 cells was constructed by using λgt10 as a vector (19–21). The library was screened with a ^{32}P-labeled *Eco*RI fragment of cDNA 14 clone (5) as described (19).

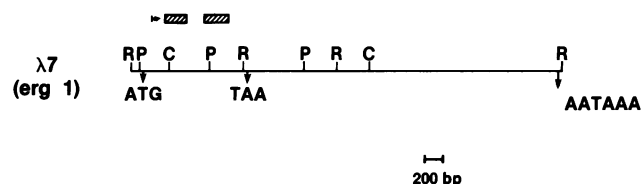FIG. 1. Analysis of the coding region of the ≈5-kb human erg1 mRNA performed on a single phage ≈4.6-kb cDNA spanning the entire region. Hatched boxes represent 5′ (left) and 3′ (right) homologous regions of v-*ets* oncogene. Initiation codon (ATG), termination codon (TAA), and polyadenylylation signal (AATAAA) are also shown. Horizontal arrow represents an 8-bp sequence, which is repeated twice as a direct repeat in another cDNA clone (erg2). Locations of restriction endonuclease sites for *Eco*RI (R), *Pvu* II (P), *Hinc* II (H), and *Cla* I (C) are indicated.

Nucleotide Sequence Analysis. The nucleotide sequence of both strands was determined by the dideoxy chain-termination method (22) after cloning individual fragments into mp18 and mp19 vectors (23). Some regions were sequenced using the method of Maxam and Gilbert (24) and also by the dideoxy chain-termination method using reverse transcriptase and synthetic primers (22).

Southern Blot Analysis. Genomic DNA (10 μg) was digested to completion with restriction enzymes, electrophoresed in a 0.8% agarose gel, and transferred to Nytran paper as suggested by the manufacturer (Fig. 2) (Schleicher & Schuell). The blots were hybridized and washed as suggested by the manufacturer.

RNA Blot Analysis. Cellular RNAs were fractionated by electrophoresis on a 1.2% agarose-formaldehyde gel (19) and transferred to Nytran filters as described above. Filters were hybridized and washed as recommended by the manufacturer (Schleicher & Schuell).

RESULTS AND DISCUSSION

Identification of an *ets*-Related Gene. A cDNA library was made from COLO 320 cells and screened with a Hu-ets2 cDNA clone (5). Three cDNA clones, ranging in size from 2.2 to 4.6 kb, were isolated. The longest cDNA clone (≈4.6 kb) was designated λ7 (Fig. 1). We named this human gene *erg* (*ets*-related gene). λ7 cDNA represents erg1, as we observed alternative splicing in another cDNA clone, which we have named erg2 (25).

The *erg* gene is probably a single copy gene, since single hybridizing fragments were detected by Southern blot analysis of genomic DNA digested separately with different restriction enzymes (Fig. 2). Comparison of different restriction enzyme patterns of genomic DNA of COLO 320 from which the cDNA library was made and of placenta DNA revealed no obvious rearrangements in the *erg* gene (Fig. 2).

---

†To whom reprint requests should be addressed.
§The recommended designation for this gene is *ETS* (7).

FIG. 2. Southern blot analysis of genomic DNAs digested with different restriction enzymes. High molecular weight DNAs from placenta (lanes 1) and COLO 320 cells (lanes 2) were digested with different restriction enzymes: (*a*) *Pst* I, (*b*) *Bam*HI, (*c*) *Bgl* II, and (*d*) *Eco*RI. They were electrophoresed in a 0.8% agarose gel, transferred to Nytran filters, and hybridized to a ^{32}P-labeled erg1 cDNA fragment (0.95-kb *Eco*RI fragment) (19). After washing under stringent conditions, hybridization was detected by autoradiography.

```
                                                                                            ***     *
GAATTCCCTCCAAAGCAAGACAAATGACTCACAGAGAAAAAAGATGGCAGAACCAAGGGCAACTAAAGCCGTCAGGTTCTGAACAGCTGGTAGATGGGCTGGCTTACTGAAGGACAT    117
**
GATTCAGACTGTCCCGGACCCAGCAGCTCATATCAAGGAACTCTCCTGATGAATGCAGTGTGGCCAAAGGCGGGAAG ATG GTG GGC AGC CCA GAC ACC GTT GGG ATG    224
                                                                            MET Val Gly Ser Pro Asp Thr Val Gly MET

      AAC TAC GGC AGC TAC ATG GAG GAG AAG CAC ATG CCA CCC CCA AAC ATG ACC ACG AAC GAG CGC AGA GTT ATC GTG CCA GCA GAT CCT ACG    314
   11 Asn Tyr Gly Ser Tyr MET Glu Glu Lys His MET Pro Pro Pro Asn MET Thr Thr Asn Glu Arg Arg Val Ile Val Pro Ala Asp Pro Thr

      CTA TGG AGT ACA GAC CAT GTG CGG CAG TGG CTG GAG TGG GCG GTG AAA GAA TAT GGC CTT CCA GAC GTC AAC ATC TTG TTA TTC CAG AAC    404
   41 Leu Trp Ser Thr Asp His Val Arg Gln Trp Leu Glu Trp Ala Val Lys Glu Tyr Gly Leu Pro Asp Val Asn Ile Leu Leu Phe Gln Asn

      ATC GAT GGG AAG GAA CTG TGC AAG ATG ACC AAG GAC GAC TTC CAG AGG CTC ACC CCC AGC TAC AAC GCC GAC ATC CTT CTC TCA CAT CTC    494
   71 Ile Asp Gly Lys Glu Leu Cys Lys MET Thr Lys Asp Asp Phe Gln Arg Leu Thr Pro Ser Tyr Asn Ala Asp Ile Leu Leu Ser His Leu

      CAC TAC CTC AGA GAG ACT CCT CTT CCA CAT TTG ACT TCA GAT GAT GTT GAT AAA GCC TTA CAA AAC TCT CCA CGG TTA ATG CAT GCT AGA    584
  101 His Tyr Leu Arg Glu Thr Pro Leu Pro His Leu Thr Ser Asp Asp Val Asp Lys Ala Leu Gln Asn Ser Pro Arg Leu MET His Ala Arg

      AAC ACA GAT TTA CCA TAT GAG CCC CCC AGG AGA TCA GCC TGG ACC GGT CAC GGC CAC CCC ACG CCC CAG TCG AAA GCT GCT CAA CCA TCT    674
  131 Asn Thr Asp Leu Pro Tyr Glu Pro Pro Arg Arg Ser Ala Trp Thr Gly His Gly His Pro Thr Pro Gln Ser Lys Ala Ala Gln Pro Ser

      CCT TCC ACA GTG CCC AAA ACT GAA GAC CAG CGT CCT CAG TTA GAT CCT TAT CAG ATT CTT GGA CCA ACA AGT AGC CGC CTT GCA AAT CCA    764
  161 Pro Ser Thr Val Pro Lys Thr Glu Asp Gln Arg Pro Gln Leu Asp Pro Tyr Gln Ile Leu Gly Pro Thr Ser Ser Arg Leu Ala Asn Pro

      GGC AGT GGC CAG ATC CAG CTT TGG CAG TTC CTC CTG GAG CTC CTG TCG GAC AGC TCC AAC TCC AGC TGC ATC ACC TGG GAA GGC ACC AAC    854
  191 Gly Ser Gly Gln Ile Gln Leu Trp Gln Phe Leu Leu Glu Leu Leu Ser Asp Ser Ser Asn Ser Ser Cys Ile Thr Trp Glu Gly Thr Asn

      GGG GAG TTC AAG ATG ACG GAT CCC GAC GAG GTG GCC CGG CGC TGG GGA GAG CGG AAG AGC AAA CCC AAC ATG AAC TAC GAT AAG CTC AGC    944
  221 Gly Glu Phe Lys MET Thr Asp Pro Asp Glu Val Ala Arg Arg Trp Gly Glu Arg Lys Ser Lys Pro Asn MET Asn Tyr Asp Lys Leu Ser

      CGC GCC CTC CGT TAC TAC TAT GAC AAG AAC ATC ATG ACC AAG GTC CAT GGG AAG CGC TAC GCC TAC AAG TTC GAC TTC CAC GGG ATC GCC    1034
  251 Arg Ala Leu Arg Tyr Tyr Tyr Asp Lys Asn Ile MET Thr Lys Val His Gly Lys Arg Tyr Ala Tyr Lys Phe Asp Phe His Gly Ile Ala

      CAG GCC CTC CAG CCC CAC CCC CCG GAG TCA TCT CTG TAC AAG TAC CCA TCA GAC CTC CCG TAC ATG GGC TCC TAT CAC GCC CAC CCA CAG    1124
  281 Gln Ala Leu Gln Pro His Pro Pro Glu Ser Ser Leu Tyr Lys Tyr Pro Ser Asp Leu Pro Tyr MET Gly Ser Tyr His Ala His Pro Gln

      AAG ATG AAC TTT GTG GCG CCC CAC CCT CCA GCC CTC CCC GTG ACA TCT TCC AGT TTT TTT GCT GCC CCA AAC CCA TAC TGG AAT TCA CCA    1214
  311 Lys MET Asn Phe Val Ala Pro His Pro Pro Ala Leu Pro Val Thr Ser Ser Ser Phe Phe Ala Ala Pro Asn Pro Tyr Trp Asn Ser Pro

                                                                                                        ***
      ACT GGG GGT ATA TAC CCC AAC ACT AGG CTC CCC ACC AGC CAT ATG CCT TCT CAT CTG GGC ACT TAC TAC TAA AGACCTGGCGGAGGCTTTTC    1306
  341 Thr Gly Gly Ile Tyr Pro Asn Thr Arg Leu Pro Thr Ser His MET Pro Ser His Leu Gly Thr Tyr Tyr **End**

      CCATCAGCGTGCATTCACCAGCCCATCGCCACAAACTCTATCGGAGAACATGAATCAAAAGTGCCTCAAGAGGAATGAAAAAAGCTTTACTGGGGCTGGGGAAGGAAGCCGGGGAAGAGA  1426
      TCCAAAGACTCTTGGGAGGGAGTTACTGAAGTCTTACTACAGAAATGAGGAGGATGCTAAAAATGTCACGAATATGGACATATCATCTGTGGACTGACCTTGTAAAAGACAGTGTATGTA  1546
      GAAGCATGAAGTCTTAAGGACAAAGTGCCAAAGAAAGTGGTCTTAAGAAATGTATAAACTTTAGAGTAGAGTTTGAATCCCACTAATGCAAACTGGGATGAAACTAAAGCAATAGAAACA  1666
      ACACAGTTTTGACCTAACATACCGTTTATAATGCCATTTTAAGGAAAACTACCTGTATTTAAAAATAGTTTCATATCAAAAACAAGAGAAAAGACACGAGAGAGACTGTGGCCCATCAAC  1786
      AGACGTTGATATGCAACTGCATGGCATGTGCTGTTTTGGTTGAAATCAAATACATTCCGTTTGATGGACAGCTGTCAGCTTTCTCAAACTGTGAAGATGACCCAAAGTTTCCAACTCCTT  1906
      TACAGTATTACCGGGACTATGAACTAAAAGGTGGGACTGAGGATGTGTATAGAGTGAGCGTGTGATTGTAGACAGAGGGGTGAAGAAGGAGGAGGAAGAGGCAGAGAAGGAGGAGACCAG  2026
      GCTGGGAAAGAAACTTCTCAAGCAATGAAGACTGGACTCAGGACATTTGGGGACTGTGTACAATGAGTTATGGAGACTCGAGGGTTCATGCAGTCAGTGTTATACCAAACCCAGTGTTAG  2146
      GAGAAAGGACACAGCGTAATGGAGAAAGGGAAGTAGTAGAATTCAGAAACAAAAATGCGCATCTCTTTCTTTGTTTGTCAAATGAAAATTTTAACTGGAATTGTCTGATATTTAAGAGAA  2266
      ACATTCAGGACCTCATCATTATGTGGGGGCTTTGTTCTCCACAGGGTCAGGTAAGAGATGGCCTTCTTGGCTGCCACAATCAGAAATCACGCAGGCATTTTGGGTAGGCGGCCTCCAGTT  2386
      TTCCTTTGAGTCGCGAAGCGCTGTGCGTTTGTCAGAATGAAGTATACAAGTCAATGTTTTTCCCCCTTTTTATATAATAATTATATAACTTATGCATTTATACACTACGAGTTGATCTCGG  2506
      CCAGCCAAAGACACACGACAAAAGAGACAATCGATATAATGTGGCCTTGAATTTTAACTCTGTATGCTTAATGTTTACAATATGAAGTTATTAGTTCTTAGAATGCAGAATGTATGTAAT  2626
      AAAATAAGCTTGGCCTAGCATGGCAAATCAGATTTATACAGGAGTCTGCATTTGCACTTTTTTTAGTGACTAAAGTTGCTTAATGAAAACATTGGCTGAATGTTGTGGATTTTGTGTTAT  2746
      AATTTACTTTGTCCAGGAACTTGTGCAAGGGAGAGCCAAGGAAATAGGATGTTTGGCACCCAAATGGCGTCAGCCTCTCCAGGTCCTTCTTGCCTCCCCTCCTGTCTTTTATTTCTAGCC  2866
      CCTTTTGGAACAGGAAGGACCCCGGGGTTTCAATTGGAGCCTCCATATTTATGCCTGGAAGGAAAGAGGCCTATGAAGCTGGGGTTGTCATTGAGAAATTCTAGTTCAGCACCTGGTCAC  2986
      AAATCACCCTTAATTCTGCTATGATTAAAATACATTTGTTGAACAGTGAACAAGCTACCACTCGTAAGGCAAACTGTATTATTACTGGCAAATAAAGCGTCATGGATAGCTGCAATTTCT  3106
      CACTTTACA [..~1.15 kb ..] CTGCAGTTAGAGGCTCACTGCTTCTCCTAAGCCTTCTGCACATGTGGCACCTGCAACCCAGGAGCAGGAGCCGGAGAGCTGCCCTCTGACAG    4379
      CAGGTGCAGCAGAGATGGCTACAGCTCAGGAGCTGGGAAGGTGATGGGGCACAGGGAAAGCACAGATGTTCTGCAGCGCCCCAAAGTGACCCATTGCCTGGAGAAAGAGAAGAAAATATT  4499
      TTTTAAAAAGCTAGTTTATTTAGCTTCTCATTAATTCATTCAAATAAAGTCGTGAGGTGACTAATTAGAGAATAAAAAATTACTTTGGACTACTCGGAATTC   4600
```

FIG. 3. The nucleotide sequence of erg1 cDNA with deduced amino acid sequence of erg1 protein. Nucleotide sequence is shown in capital letters. Amino acid sequence predicted from the long open reading frame and designated by the three-letter code is shown below the nucleotide sequence. Nucleotide positions are indicated on the right; amino acid positions are indicated on the left. The three termination codons in-frame with the reading frame upstream and downstream are indicated by asterisks. The 8-bp sequence, which is repeated twice in another cDNA (25), is shown by a wavy line. Approximately 1.15 kb of the 3' noncoding region that has not been sequenced is shown in brackets. The two potential polyadenylylation signals are double underlined.
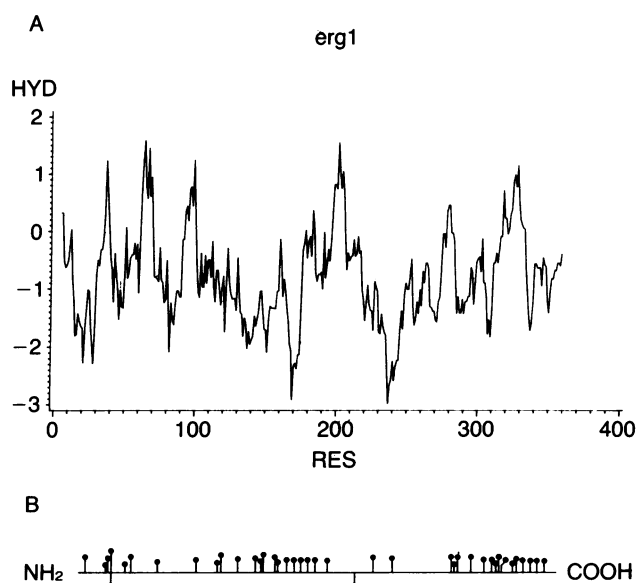
A

erg1

HYD

FIG. 4. (*A*) Hydropathicity plot of the deduced erg1 protein. Hydropathicity values were calculated as described by Kyte and Doolittle (31). Hydrophobic values are positive. (*B*) Position of proline residues (●) and potential N-linked glycosylation sites (forks below line) in the predicted erg1 protein.

Nucleotide and Predicted Amino Acid Sequence of erg1. The nucleotide sequence of the λ7 cDNA clone (erg1) was determined by a combination of the dideoxy chain-termination method of Sanger and also by the method of Maxam and Gilbert. Analysis of the λ7 cDNA clone revealed that it is ≈4600 base pairs (bp) long and contained a complete open reading frame (Fig. 3). The longest open reading frame starting with a methionine codon at position 195 in the nucleotide sequence encodes a 363 amino acid polypeptide (Fig. 3). There are five methionine codons present close to each other at the 5′ end (Fig. 3) and it is not certain which of these methionines is being used. Although the flanking nucleotides of the predicted initiation codon ATG do not show a perfect match with Kozak's consensus sequence

[CC(A/G)CCATGG] (26, 27), this codon (at position 195) seems to be the initiation codon because two stop codons are found in the reading frame upstream in the 5′-untranslated region (Fig. 3). The only long open reading frame with an initiation codon (position 195) is preceded by another ATG in the same reading frame. This occurs upstream of the termination codon that precedes the long open reading frame. Thus, a polypeptide (28 amino acids) could also be potentially synthesized from erg1 mRNA in the bicistronic fashion proposed for certain other eukaryotic mRNAs (27–30). However, the context of the ATG at position 24 is less favorable than the context of the ATG at position 195 as the former ATG lacks a purine in position −3, a feature that is believed important for initiation of translation (26, 27). Thus, erg1 mRNA joins the growing list of exceptions to the proposal that eukaryotic mRNAs initiate at the first AUG nearest the 5′ end (26, 27). The extreme 3′ sequence must have been lost from the erg1 clone during the cloning of cDNA because of the absence of the poly(A) sequence at the 3′-untranslated region. However, we could see two potential polyadenylylation signals in the AATAAA sequence at the 3′ terminus (Fig. 3). It is uncertain which of the two is used as a polyadenylylation signal. The other structural feature is the presence of an 8-bp sequence (Figs. 1 and 3), which is repeated twice as a direct repeat in another cDNA clone, erg2 (25).

The deduced amino acid sequence of the erg1 polypeptide shows that it has a primary length of 363 amino acids, is proline rich (41 residues) (Figs. 3 and 4*B*), and contains 13 methionines. Interestingly, prolines are clustered in the middle and carboxyl-terminal regions of erg1 protein (Fig. 4*B*). It contains two potential glycosylation sites at amino acid positions 25 and 210. In general, the erg1 polypeptide appears to be hydrophilic in nature (Fig. 4*A*). *In vitro* transcription and translation of erg1 cDNA supported the expected molecular mass (≈41 kDa) of the presumed encoded protein obtained from nucleotide sequence analysis (25).

A computer analysis of the nucleotide sequence revealed that the 3′ (nucleotides 763–1025) and 5′ (nucleotides 305–482) regions of erg1 showed ≈70% and ≈56% homology with the 3′ (nucleotides 1665–1927) and 5′ (nucleotides 877–1051) domains of the v-*ets* oncogene, respectively. Comparison of the amino acid sequences specified by erg1 to those specified by v-*ets* and Hu-*ets2* revealed 70–74% homology over a
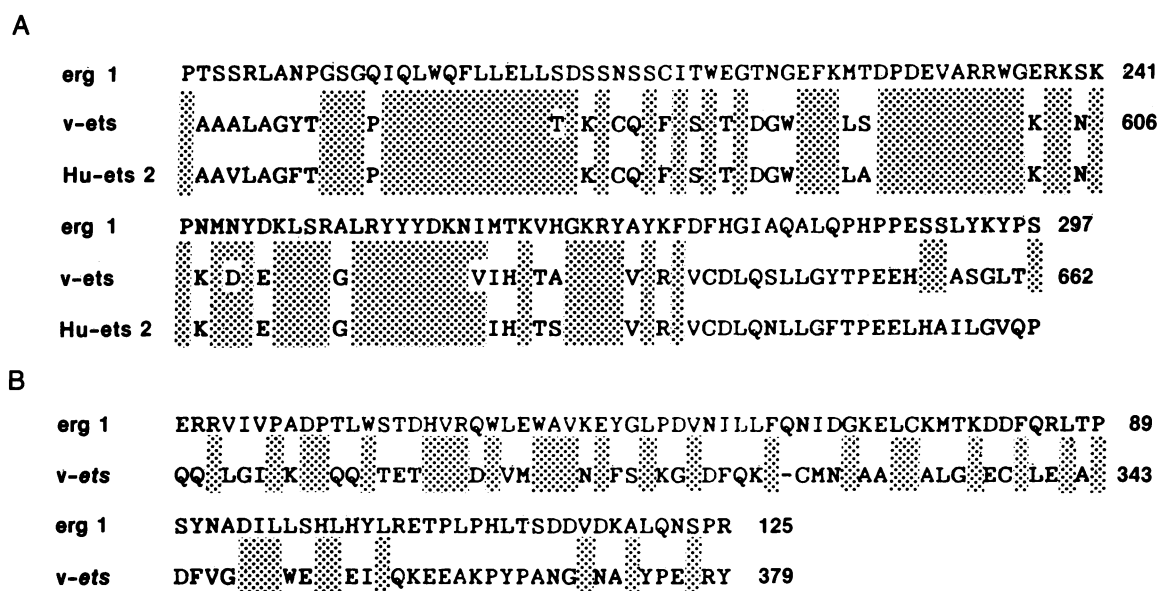
A

B

FIG. 5. Comparison of the nucleotide and the predicted amino acid sequences of (*A*) 3′ (upper) and (*B*) 5′ (lower) regions of *erg1* with v-*ets* (3) and Hu-*ets2* (5). Stippled regions represent identical amino acids. Single-letter abbreviations for the amino acid residues are used. The numbering system given for v-*ets* is according to Nunn *et al.* (3).
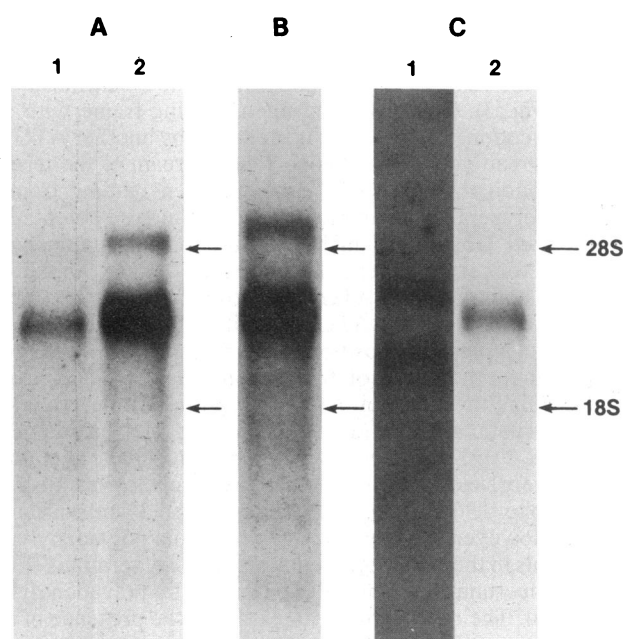
A   B   C

1   2   1   2

FIG. 6.   RNA blot analysis of poly(A)$^+$ RNA (10 μg) isolated from COLO 320 and Molt-4 cells. (*A*) Polyadenylylated RNA from Molt-4 (lane 1) and COLO 320 (lane 2) cells was analyzed by denaturing gel electrophoresis in formaldehyde (19), transferred to Nytran filters, and hybridized with ^{32}P-labeled erg1 probe as described in Fig. 1. Sizes were determined by reference to rRNA and Bethesda Research Laboratory RNA ladder standard markers detected by staining the gel with ethidium bromide. (*B*) Longer exposure of RNA blot analysis of poly(A)$^+$ RNA from COLO 320 cells hybridized to erg1 probe. (*C*) Hybridization of erg1 probe to poly(A)$^+$ RNA from COLO 320 cells (lane 2). The same blot is washed and reprobed with Hu-ets2 cDNA probe (lane 1).

region of 84 amino acids (positions 191–274) (Fig. 5*A*). The similarity is even more striking when one considers that at least 10 of the 26 amino acid differences are conservative. Taking this into account, the homology in this region is 82–85%. In addition, the homology region extends 77 amino acids further toward the amino-terminal region (positions 114–191). It was shown previously that Hu-*ets2* has no homology to the 5' region of the v-*ets* oncogene (5). However, *erg1* shows homology of 40% over a region of 72 amino acids (positions 32–103) (Fig. 5). The homology in this region becomes striking when one considers that at least 17 of the 43 amino acid differences are conservative. Taking this into account, the homology in this region is 64%. Thus, *erg1* shares homology to two domains of the v-*ets* oncogene, suggesting that they each share a set of biochemical functions. Furthermore, the carboxyl end of the *erg1* gene product appears to be long, relative to the v-*ets* and Hu-*ets2* gene products. These additional amino acids suggest the unique nature of this gene and possibly specify specialized function(s) of the erg1 protein.

A computer-assisted search of the National Biomedical Research Foundation protein data base and comparison of the amino acid sequences specified by *erg1* exhibited 40% homology with Herpes simplex virus thymidine kinase over a range of 47 amino acids. erg1 also exhibited a low homology with the polymerase polyprotein of human immunodeficiency virus (HIV-1) and Rous sarcoma virus over a range of 310 amino acids. However, a homology of 24% with HIV-1 is seen over a range of 62 amino acids near the carboxl-terminal region of erg1 protein (positions 243–305). Thirty amino acids near the amino-terminal region of erg1 protein also exhibited 39% homology with the genome polyprotein of foot and mouth disease virus. Proline-rich phosphoproteins (human),

gag protein of Abelson, Moloney murine leukemia virus, cAMP-dependent protein kinase, abl transforming protein, and Thy-1 membrane glycoprotein precursor (rat) also exhibited 20–30% homology with *erg1*. The significance of these homologies, if any, remains to be established.

RNA Blot Analysis of *erg* Gene Transcripts. The nucleotide and amino acid sequences of erg1 indicated that the *erg* gene is related to the v-*ets* oncogene and Hu-*ets2* gene (Fig. 5). It was shown that Hu-*ets2* is expressed in COLO cells as 4.7-, 3.2-, and 2.7-kb transcripts (5). To determine the length of the *erg* gene transcripts, we hybridized size-fractionated poly(A) mRNA from COLO 320 and Molt-4 cells with *erg*-specific probe (0.95-kb *Eco*RI fragment of erg1 clone, which does not share homology with Hu-ets2 cDNA). Major transcripts ranging in size from 3.2 to 3.6 kb and minor transcripts of ≈5 kb were detected (Fig. 6). It appears that these transcripts may have been generated by alternative splicing and/or polyadenylylation (25). On longer exposure, minor transcripts of smaller size were seen (Fig. 6*B*). It is not certain whether these transcripts are real or degradation products. The same blot was washed and reprobed with Hu-ets2 cDNA probe. As shown in Fig. 6*C*, the sizes of the transcripts of the *erg* gene and the Hu-*ets2* gene were different, providing additional evidence for the *erg* gene being distinct from the Hu-*ets2* and Hu-*ets1* genes.

Implications. Characterization of another cDNA clone (erg2) revealed an alternative initiation, alternative splicing, and alternative polyadenylylation, giving rise to another ets-related polypeptide (25). Thus, the *erg* gene appears to encode for at least two proteins that share a limited domain(s) of homology with v-*ets* and Hu-*ets2*-encoded protein products. Recently, evidence has been provided for the existence of a set of proteins in chicken related to, but distinct from, the chicken c-*ets*-encoded protein p54, supporting the presence of ets-related proteins (32). More extensive analysis of these similar but distinct genes, including *in situ* analysis of mRNAs and proteins, is required to determine whether these products are specific to a certain cell lineage or whether they play a more general role in cell growth.

A number of genes related to but distinct from oncogenes transduced by retroviruses have been implicated in the neoplastic process. For example, N-*myc* (8, 9), L-*myc* (10), and c-*erb* B2 (11) were shown to be amplified in neuroblastomas, small lung cell carcinomas, and human mammary carcinomas, respectively. Interestingly, it was shown that the Hu-*ets1* and the Hu-*ets2* genes are translocated in certain leukemias (33, 34). It remains to be seen whether *erg*, a gene related to the *ets* oncogene, can be linked to any human malignancy either by amplification, rearrangement, or translocation.

1.   Moscovici, C., Samurat, T., Gazzolo, L. & Moscovici, M. G. (1981) *Virology* **111**, 765–768.
2.   Radke, K., Beug, H., Kornfeld, S. & Graf, T. (1982) *Cell* **31**, 643–653.
3.   Nunn, M. F., Seeburg, P. H., Moscovici, C. & Duesberg, P. H. (1983) *Nature (London)* **306**, 391–395.
4.   Leprince, D., Gegonne, A., Coll, J., De Tainse, C., Schneeberger, A., Lagrou, C. & Stehelin, D. (1983) *Nature (London)* **306**, 395–397.
5.   Watson, D. K., McWilliams-Smith, M. J., Nunn, M. F., Duesberg, P. H., O'Brien, S. J. & Papas, T. S. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 7294–7298.
6.   Watson, D. K., McWilliams-Smith, M. J., Kozak, C., Reeves, R., Gearhart, J., Nash, W., Modi, W., Duesberg, P., Papas, T. S. & O'Brien, S. J. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 1792–1796.

Biochemistry: Reddy *et al.*

Proc. Natl. Acad. Sci. USA 84 (1987)   6135

7. Eighth International Workshop on Human Gene Mapping (1985) *Cytogenet. Cell Genet.* **40** (1–4).
8. Schwab, M., Alitalo, K., Klempnauer, K.-H., Varmus, H., Bishop, J. M., Gilbert, F., Brodeur, G., Goldstein, M. & Trent, J. (1983) *Nature (London)* **305**, 245–248.
9. Kohl, N. E., Kanda, N., Schreck, R. R., Burns, G., Latt, S. A., Gilbert, F. & Alt, F. W. (1983) *Cell* **35**, 359–367.
10. Nau, M., Brooks, B., Battey, J., Sansbille, E., Gazdar, A., Kirsch, I., McBridge, O. W., Bertness, V., Hollis, G. & Minna, J. (1985) *Nature (London)* **318**, 69–73.
11. King, C. P., Kraus, M. H. & Aaronson, S. A. (1985) *Science* **229**, 974–976.
12. Bergmann, C. I., Hung, M. & Weinberg, R. (1986) *Nature (London)* **319**, 226–230.
13. Kruth, G. D., King, C. R., Kraus, M. H., Popescu, N. C., Amsbaugh, S. C., McBride, W. O. & Aaronson, S. A. (1986) *Science* **234**, 1545–1548.
14. Kawakami, T., Pennington, C. Y. & Robbins, K. C. (1986) *Mol. Cell. Biol.* **6**, 4195–4201.
15. Semba, K., Noshizawa, M., Miyazima, N., Yoshida, M. C., Suekegawa, J., Yamanashi, T., Sasaki, M., Yamato, T. & Toyoshima, K. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 5459–5463.
16. Mark, G. E., Seeley, T. W., Shows, T. B. & Mountz, J. D. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 6312–6316.
17. Chardin, P. & Tavitian, A. (1986) *EMBO J.* **5**, 2203–2208.
18. Madaule, P. & Axel, R. (1985) *Cell* **41**, 31–40.
19. Maniatis, T., Fritsch, F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
20. Huynh, T. V., Young, R. A. & Davis, R. W. (1985) in *DNA Cloning*, ed. Glover, D. M. (IRL, Oxford), Vol. 1, pp. 49–78.
21. Gubler, U. & Hoffman, B. J. (1983) *Gene* **5**, 263–269.
22. Sanger, F. S., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
23. Messing, J. (1983) *Methods Enzymol.* **101**, 20–78.
24. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–559.
25. Rao, V. N., Papas, T. S. & Reddy, E. S. P. (1987) *Science*, in press.
26. Kozak, M. (1983) *Microbiol. Rev.* **47**, 1–45.
27. Kozak, M. (1984) *Nucleic Acids Res.* **12**, 3873–3893.
28. Hughes, S., Mellestrom, K., Kosik, E., Tamanoi, F. & Brugge, J. (1984) *Mol. Cell. Biol.* **4**, 1738–1746.
29. Liu, C. C., Simonsen, C. C. & Levinson, A. D. (1984) *Nature (London)* **309**, 82–85.
30. Mardon, G. & Varmus, H. E. (1983) *Cell* **32**, 871–879.
31. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132.
32. Ghysdael, J., Gegonne, A., Pognonec, P., Boulukos, K., Leprince, D., Dernis, D., Lagrou, C. & Stehelin, D. (1986) *EMBO J.* **5**, 2251–2256.
33. Diag, M. O., Le Beau, M. M., Pitha, P. S. & Rowley, J. D. (1986) *Science* **231**, 265–268.
34. Sacchi, N., Watson, D. K., Guerts Van Kessel, A. H. M., Hagemeijer, A., Kersey, J., Drabkin, H. D., Patterson, D. & Papas, T. S. (1986) *Science* **231**, 379–381.